DOCUMENT RESUME

ED 039 033 **PS 003 001**

AUTHOR	Espinosa, Renato; Pierce-Jones, John
TITLE	A Report on the Results of the Administration of the Gumpgookies Test to the Texas Evaluation Sample. Part of the Final Report on Head Start Evaluation and Research: 1968-69 to the Office of Economic Opportunity.
INSTITUTION	Texas Univ., Austin. Child Development Evaluation and Research Center.
SPONS AGENCY	Office of Economic Opportunity, Washington, D.C.
REPORT NO	OEO-4115
PUB DATE	Aug 69
NOTE	33p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.75
DESCRIPTORS	Anglo Americans, Item Analysis, Mexican Americans, *Motivation, Negroes, Parent Attitudes, Preschool Children, *Test Construction, Test Interpretation, Test Reliability, *Visual Measures
IDENTIFIERS	Gumpgookies, Head Start

ABSTRACT

An experimental test for achievement motivation which consists of 100 pictures of imaginary figures called gumpgookies was tested on Head Start children. On each picture, two gumpgookies are presented in a semi-structured situation, each engaged in behaviors that reflect different degrees of motivation to achieve. The child being tested is asked to select the gumpgookie on each picture which is most like himself, doing what he would do. Subjects were 179 Negro, Mexican-American, and Anglo children. Parent interview data were examined and subscales were constructed to assess educational opportunities, aspirations and attitudes, physical control, rejection, guilt, and reaction to infraction. The 100 items of the test were classified depending on (1) verbal clues alone, (2) visual clues alone, (3) both verbal and visual clues; and were also classified as either short or long and easy or hard. An item analysis indicated that certain subjects responded positionally throughout the test. Some items were eliminated because of lack of response consistency on a retest. Although positional preference problems with the version of the test used in this study have necessitated major alterations, a new version of the test has been developed which should significantly decrease the positional response set. (MH)

PART OF THE FINAL REPORT

ON

HEAD START EVALUATION AND RESEARCH: 1968-69

TO

THE OFFICE OF ECONOMIC OPPORTUNITY
(Contract No. OEO-4115)

CHILD DEVELOPMENT EVALUATION AND RESEARCH CENTER

John Pierce-Jones, Ph.D., Director

The University of Texas at Austin

The University of Texas at Austin

August, 1969

A REPORT ON THE RESULTS OF THE ADMINISTRATION OF THE GUMPGOOKIES TEST TO THE TEXAS EVALUATION SAMPLE

Renato Espinosa
John Pierce-Jones

This study was supported by Contract No. OEO-4115 between the Office of Economic Opportunity, Project Head Start, and The University of Texas at Austin.

A Report on the Results of the Administration of the Gumpgookies Test to the Texas Evaluation Sample

Renato Espinosa

John Pierce-Jones

The University of Texas at Austin

Background

As part of the 1968-69 Evaluation of Head Start, Gumpgookies, an experimental test for Achievement Motivation was administered. This instrument is being developed at the University of Hawaii's Head Start Evaluation Center by Drs. Dorothy C. Adkins and Bonnie Ballif.

The Gumpgookies version administered for pretest consists of 100 pictures of imaginary figures called Gumpgookies. On each picture, two Gumpgookies are presented in a semi-structured situation, each engaged in behaviors that reflect different degrees of motivation. The dichotomous options are designed to reflect the strength of the learned responses hypothesized to be constituents of motivation to achieve. The child is told that he has his own Gumpgookie that follows him, does what he does, likes what he likes, etc. The subject's task is to identify and point to his own Gumpgookie.

The type of response (choice between two alternatives) raises the problem of random answers, particularly when the choice

is a difficult one to make because of similarity between the two stimuli, difficulty of the concepts involved, length, etc. Any factor responsible for random answers could possibly have an effect on the test-retest reliability of the instrument.

It was hypothesized that the type of discriminative clues, the length of the text and the difficulty level of the concepts might affect the response consistency over a short period of time. At the same time, subjects could possibly choose on the basis of position preference independent from the actual content of the item.

An examination of the items from the point of view of the discrimination task involved, independently of the content of the item, revealed three types of items:

a) Items in which the choice had to be made on the basis of Verbal clues alone, in which the two Gumpgookies look alike but are described by the tester as being different;

b) Items in which the two Gumpgookies are engaged in the same activity and the subject is directed to choose one on the basis of Visual clues alone; and

c) Items in which both the verbal description and the picture of the Gumpgookies provide the basis for choosing between them.

Taking into account the total length of the text, including introduction and description of the alternatives, the items

can be classified into:

a) short items (20 words or less) and

b) long items (21 words or less).

It is assumed that if the text for an item is too long the subject might forget the first part and then choose on the basis of some position bias.

In terms of the difficulty of the items, they can be classified into:

a) low difficulty level and

b) high difficulty level.

It is assumed that if a subject cannot recognize the situation presented, or any of the concepts involved, he would tend to choose at random or on the basis of some positional preference.

It was suspected, then, that whenever the discrimination between the two Gumpgookies was a difficult one because of lack of familiarity with the content, situation, expressions, wording, etc., children would tend to choose at random. Moreover, it is possible that some subjects would have some preference for either right or left side (position preference) or some preference for the first Gumpgookie described (primacy) or the last one pointed to (recency).

This report presents an updated and more complete description of a check on response consistency that was performed with the pretest data and sent to Drs. Adkins and Datta. Also, we will present some additional analyses performed with the posttest data.

The Sample

The number of children used for the different analyses vary somewhat due to the unavailability of data for some instruments. In general, the subjects were Negro, Mexican-American and Anglo children enrolled in Head Start in Austin, Texas during the 1968-69 period.

Pretest Data

Descriptive Statistics

The 100 Item version of the Gumpgookies was administered to a total of 179 subjects at pretest time. Table 1 presents the age in months at pretest time for the total sample classified in terms of ethnic group and sex. The analysis of variance revealed that there are significant differences in age for the three ethnic groups, Anglo children being the oldest group and Negro children the youngest.

Table 2 presents the analysis of variance of scores in the Gumpgookies Test at pretest time for the sample classified according to ethnic group and sex. There are significant differences for both ethnic group and sex. Ethnic groups are ordered from Mexican-Americans (lowest) to Anglos (highest). Females scored higher than males.

In order to explore further the ethnic differences found in the pretest scores, the Parent Interview data was examined. Seven subscales were constructed with items thought to be relevant to Achievement Motivation.

TABLE 1

Ethnic Group, Sex, N and Age in Months

Total Sample N = 179

Ethnic Group	Sex	N	Mean	S. D.
Negro	Males	49	66.78	4.00
Negro	Females	32	66.53	4.06
Mexican-American	Males	34	67.38	3.68
Mexican-American	Females	48	68.73	2.98
Anglo	Males	8	69.88	3.56
Anglo	Females	8	70.38	3.20

TABLE 2

Analysis of Variance of Achievement Scores

by Ethnic Group and Sex

Source	M.S.	D.F.	F	P	Var.
Total	123.169	178			
Between	427.134	5			
A (Ethnic)	872.058	2	7.6239	.0010	.0688
B (Sex)	386.691	1	3.3806	.0642	.0124
A X B	2.431	2	.0213	.9797	0.
Within	114.384	173			

Means for all Effects

	1 Negro	2 Mex.-Am.	3 Anglo
A Main Effect Ethnic Group	74.41	67.24	77.00

	1 Male	2 Female
B Main Effect Sex	70.94	74.83

A X B Interaction	Sex 1	 2
Ethnic Group		
1	72.76	76.06
2	65.06	69.42
3	75.00	79.00

TABLE 2 (Continued)

Analysis of Variance of Achievement Scores

by Ethnic Group and Sex

Cell	Number	Means	S. D.
A1 B1	49	72.76	11.22
A1 B2	32	76.06	11.53
A2 B1	34	65.06	9.88
A2 B2	48	69.42	10.74
A3 B1	8	75.00	7.86
A3 B2	8	79.00	8.99

Subscale 1. Educational Opportunities. Responses to items 15, 15a, 15b, 15c, 16 and 17 were added to get a score. A high score indicates more and better educational opportunities at home.

Subscale 2. Educational Aspirations for the Child. Items 18, 19, 20, 21 and 22 were combined to form a scale. A high score indicates a high level of aspiration for the child. Item 22 was reversed.

Subscale 3. Educational Attitudes. The responses to items 23b, i, k, l, n and t were combined with items k, n, and t reversed. A high score indicates a positive attitude towards education.

Subscale 4. Physical Control. Responses to items 26a and 27a were reversed and combined to obtain a score. A high score indicates avoidance of physical punishment.

Subscale 5. Psychological Control - Rejection. Responses to items 26b and 27b were reversed and combined. A high score indicates absence of rejection as a technique of control.

Subscale 6. Psychological Control - Guilt. Responses to items 26c and 27c were reversed and combined. A high score indicates absence of use of guilt as a technique for control.

Subscale 7. Reaction to Infraction. Responses to items 26d and 27d were added to form a scale. A high score indicated a more constructive response to infraction.

Total Score, the simple addition of all seven subscales, indicates a more positive, rewarding environment.

Table 3 presents the results of the analysis of variance of the subscales and total score for subjects classified by ethnic group and level of Achievement Motivation (median split).

In general, several scales show significant differences between ethnic groups in favor of Negro subjects. This is consistent with the differences in achievement scores found previously. However, for subjects classified by level of Achievement Motivation measured with the Gumpgookies Test only Scale 2, Educational Aspirations, shows a significant difference in favor of High Achievement subjects. This is, in general, consistent with the theoretical foundations of the test, indicating a family environment that sets higher educational goals. The data from the Parent Interview, however, can be easily contaminated by social desirability and interviewer demands.

Next, individual scores were analyzed in terms of frequency of right and left choices. Also, since a score of 50 can be obtained by answering at random, a count was made of the number of subjects obtaining a score not significantly greater than chance (less than 60).

1. Out of a total of 179 subjects, 33 (18%) obtained scores of 60 or less, scores that are not significantly greater than a chance score of 50.
2. Of these 33 subjects, 23 showed a significant "positional response set" (either right or left), while the remaining 10 subjects seemed to be responding at random.

TABLE 3

Two-way Analysis of Variance of Seven Subscales and Total Score on Responses to the Pretest Parent Interview for Two Ethnic Groups Classified into High and Low (Median Split) Achievement Scores on the 100 Item Pretest Gumpgookies Test (N = 118)

Scale	Negro Mean	Mex.-Am. Mean	p	Low Achievement Mean	High Achievement Mean	p
Scale 1	14.19	12.52	.04	13.05	13.66	n.s.
Scale 2	26.87	23.33	.002	23.66	26.54	.01
Scale 3	29.92	30.45	n.s.	30.50	29.88	n.s.
Scale 4	9.88	9.45	n.s.	9.79	9.54	n.s.
Scale 5	13.25	12.47	.09	12.90	12.82	n.s.
Scale 6	13.42	12.56	.06	13.07	12.91	n.s.
Scale 7	5.10	4.44	.08	4.80	4.74	n.s.
Total Score	112.63	105.22	.003	107.77	110.09	n.s.

3. In addition, 27 subjects showed a significant "positional response set" although their score was significantly greater than the chance score of 50. Thus, a total of 50 subjects (27%) showed "positional response set"; 28 subjects showed a preference for the right hand picture and 22 subjects showed a preference for the left hand picture.

The incidence of positional response bias (defined as 60 or more choices of either right or left) was analyzed by ethnic group, sex, and age. The results are presented in Table 4.

Item Analysis of the Scale

An item analysis was performed for the 100 items using the total sample of 179 subjects.

The Alpha Coefficient of Internal Consistency was found to be .8622. The mean Achievement Score was 71.37 with a Standard Deviation of 11.26.

Table 5 presents the Item-Scale correlations for 100 items while Table 6 presents the items classified according to the size of the Item-Scale correlations and significance levels.

There are four items that correlate negatively (nonsignificantly) with the scale: Nos. 4, 8, 9, and 61.

There are eight items which do not correlate significantly with the scale: Nos. 5, 6, 24, 36, 46, 50, 56, and 64.

TABLE 4

Incidence of Positional Response Bias for Two Age Groups*

	Male	Female	Total
	Younger Subjects		
Negro	9	7	16
Mexican-American	5	5	10
Anglo	1	1	2
Total	15	13	28
	Older Subjects		
Negro	6	1	7
Mexican-American	5	10	15
Total	11	11	22

*Each subgroup (i.e., Negro-Males) was divided at their median age to generate younger and older subject groups.

TABLE 5

Item-Scale Correlations for 100 Items (N = 179)

Item	r.	Item	r.	Item	r.	Item	r.
1	.31	26	.23	51	.25	76	.31
2	.20	27	.31	52	.38	77	.34
3	.26	28	.19	53	.27	78	.31
4	-.03	29	.38	54	.17	79	.30
5	.12	30	.18	55	.27	80	.42
6	.12	31	.36	56	.14	81	.24
7	.31	32	.37	57	.27	82	.39
8	-.12	33	.33	58	.47	83	.25
9	-.03	34	.16	59	.39	84	.33
10	.25	35	.31	60	.41	85	.23
11	.29	36	.13	61	-.05	86	.32
12	.24	37	.23	62	.32	87	.16
13	.25	38	.30	63	.40	88	.37
14	.22	39	.30	64	.01	89	.18
15	.34	40	.16	65	.28	90	.23
16	.28	41	.41	66	.39	91	.28
17	.27	42	.38	67	.24	92	.37
18	.43	43	.42	68	.34	93	.28
19	.22	44	.31	69	.36	94	.34
20	.27	45	.44	70	.17	95	.24
21	.27	46	.14	71	.35	96	.24
22	.25	47	.29	72	.26	97	.45
23	.42	48	.25	73	.27	98	.37
24	.14	49	.15	74	.37	99	.27
25	.23	50	.11	75	.37	100	.40

TABLE 6

Item-Scale Correlations and Significance Levels

for 100 Items of the Gumpgookies (N = 179)

Correlation	p	Item No.
.4000 to .4650	.000	18 , 23 , 41 , 43 , 45 , 58 , 60 , 80 , 97 ,100 .
.3000 to .3999	.000	1 , 7 , 15 , 27 , 29 , 31 , 32 , 33 , 35 , 38 , 39 , 42 , 44 , 52 , 59 , 62 , 63 , 66 , 68 , 69 , 71 , 74 , 75 , 76 , 77 , 78 , 79 , 82 , 84 , 86 , 88 , 92 , 94 , 98 .
.2000 to .2999	.001	3 , 10 , 11 , 12 , 13 , 14 , 16 , 17 , 19 , 20 , 21 , 22 , 25 , 26 , 37 , 47 , 48 , 51 , 53 , 55 , 57 , 65 , 67 , 72 , 73 , 81 , 83 , 85 , 90 , 91 , 93 , 95 , 96 , 99 .
.1500 to .1999	.05	2 , 28 , 30 , 34 , 40 , 49 , 54 , 70 , 87 , 89 .
.0000 to .1499	n.s.	5 , 6 , 24 , 36 , 46 , 50 , 56 , 64 .
negative...	n.s.	4 , 8 , 9 , 61 .

There are ten items that correlate low but significantly with the scale: Nos. 2, 28, 30, 34, 40, 49, 54, 70, 87, and 89.

The remaining items correlate significantly with the scale at P = .01 or better.

The analysis indicates that there are 22 items that are not satisfactory. This is a purely internal analysis without any reference to external criteria of validity. In considering modifications of the scale this is not the only consideration that is important.

A Check on the Response Consistency (test-retest)

Procedure

On the basis of the factors mentioned in the background section of this report, 10 items were selected to represent the combination of factors thought to be important in determining test-retest consistency. The ten selected items are presented in Table 7. Because of the suspected position response bias, the drawings were reversed, as well as the description of the alternatives. The ten items were then administered at the end of the regular testing session. Thus, a comparison can be made between a subject's response to the same item but with the alternatives read in different sequences. If the choice of one alternative is determined by a preference for motivated behavior, we would expect that subject to choose the same alternative again, although the second time it is read in a different order as before. If, on the other hand a subject chose the motivated Gumpgookie because it was read

TABLE 7

Ten Selected Items Readministered at the End of the Regular Testing Session

Type of Discrimination	Length	Difficulty	Item Number
Verbal clues	Short	Low	4
		High	28
	Long	Low	8
		High	25
Visual and Verbal clues	Short	Low	41
		High	81
	Long	Low	68
		High	52
Visual clues	Short	Low	34
		High	84

first (or last) or because of position preference, the second time he would tend to choose the same side, but this time his response is scored as not motivated (wrong).

The additional 10 items were administered to only 159 subjects out of the total pretest sample of 179. The decrease in number was due to failure of two testers to administer the additional items. The subjects deleted, however, are a random subsample.

Considering the subject's response to the same item twice, his responses can be classified either "consistent" or "inconsistent." One can be consistent by choosing the "correct" or "incorrect" response both times. On the other hand, inconsistent subjects changed their response from "correct" to "incorrect" or vice-versa.

Table 8 presents the data for the ten items that were readministered. "C" is "correct," "I" is "incorrect," "T" is total. The first letter corresponds to the first administration and the second to the readministration.

Inspection of Table 4 reveals the following facts about the responses to the ten selected items that were readministered:

1. Item No. 4 (101). About 50% of the subjects changed their responses on the second administration. Of that 50%, the changes that occurred were divided about equal between Correct-Incorrect (C-I) and Incorrect-Correct (I-C) on the first and second administration respectively. Since the item

TABLE 8

Analysis of Answers to Repeated Administration

of Ten Selected Items (N = 159)

Item Number		Consistent			Inconsistent		
Adm.	Readm.	C-C	I-I	T	C-I	I-C	T
4	101	36	43	79	47	33	80
8	102	26	40	66	22	71	93
25	103	88	10	98	36	25	61
28	104	29	51	80	37	42	79
34	105	96	23	119	19	21	40
41	106	102	19	121	21	17	38
52	107	119	4	123	20	16	36
68	108	108	10	118	23	18	41
81	109	23	20	43	81	35	116
84	110	96	14	110	13	36	49

was reversed for the second administration, the changes can be explained as the result of a tendency to choose the same position the second time ("positional response set").

2. Item No. 8 (102). About 40% of the subjects responded consistently to both administrations, while significantly more of the subjects who changed did so from Incorrect to Correct the second time.
3. Item No. 25 (103). About 60% of the subjects responded consistently to these items. Those who changed their responses on the second administration are almost equally divided between Correct-Incorrect and Incorrect-Correct.
4. Item No. 28 (104). About 50% of the subjects changed responses on the second administration and are equally divided between Correct-Incorrect and Incorrect-Correct.
5. Items No. 34 (105), 41 (106), 52 (107) and 68 (108). About 75% of the subjects responded consistently to these items. Those who changed are equally divided among Correct-Incorrect and Incorrect-Correct.
6. Item No. 81 (109). Only about 25% of the subjects responded consistently to this item. Of those who changed, a significantly greater proportion changed from Correct to Incorrect.
7. Item No. 84 (110). About 70% of the subjects responded consistently to this item in both administrations. Of those who changed, significantly more subjects did so from Incorrect to Correct.

In general, it can be said that those items which were answered consistently in both administrations are "good" items, regardless of the absolute number of correct responses (e.g., items 25, 34, 41, 52, 68, and 84). When the percentage of consistent and inconsistent responses is about equal (e.g., items 4, 28) and the reversals are not systematic, then random answers will have to be suspected.

There are two items (4, 81) where there are more inconsistent than consistent answers. For item 8, the greater majority of subjects changed from Incorrect to Correct. This suggests that the subjects who were not sure of their answers tended to repeat the same answer the second time due to a "positional response set."

If we examine "good" and "bad" items according to our classification of the discrimination task involved, we see that, as expected, three of the four "bad" items are items where the discrimination task is in terms of Verbal clues alone with no visual differences between the Gumpgookies involved. The remaining "bad" item has both Visual and Verbal clues, but a High Level of Difficulty due to the use of the concepts "king" versus "leader." The "good" items in terms of consistency were found to correlate with the scale from .19 to .45. Of the four "bad" items, two (4 and 8) correlate nonsignificantly with the scale, while items 28 and 81 correlate significantly. These data are presented in Table 9.

TABLE 9

Characteristics of the Ten Items

Repeated at the End of the Scale

Type of Discrimination	Length	Difficulty	Item Number	
Verbal clues	Short	Low	4	Bad Item
		High	28	Bad Item
	Long	Low	8	Bad Item
		High	25	Good Item
Visual and Verbal clues	Short	Low	41	Good Item
		High	81	Bad Item
	Long	Low	68	Good Item
		High	52	Good Item
Visual clues	Short	Low	34	Good Item
		High	84	Good Item

Table 10 presents a classification of those items found to correlate nonsignificantly or negatively with the scale and the type of discrimination they involve. It can be seen that 11 out of 12 items call for discriminations on the basis of Verbal clues alone. Although not all the Verbal discrimination type items are "bad" items, this clearly indicates that for children of this age a purely "projective" test will elicit mainly random answers or positional or temporal response sets.

Table 11 examines the type of discrimination called for by the ten best items in terms of correlation with the scale. Table 11 reveals that 8 out of 10 "good" items offer to the subject both Visual and Verbal clues. The two remaining items are both short, and low in difficulty level.

The objectivity of our classification can be questioned, particularly the variables of length and difficulty level. The difficulty level was determined as an educated guess based on past experience testing Head Start children. The type of discrimination can be classified with enough objectivity as to be useful.

Posttest Data

At posttest time a shortened version of the Gumpgookies was administered. This new version consisted of 55 items and the items eliminated were those reported by most Evaluation Centers as being "bad" items. The items deleted had in general low Item-Scale correlations in our sample.

TABLE 10

Discrimination Type and Item-Scale Correlations

Item	r.	Discrimination	Length	Difficulty
4	-.03	Verbal	Short	Low
5	.12	Verbal	Long	Low
6	.11	Verbal	Short	High
8	-.12	Verbal	Short	Low
9	-.03	Verbal	Long	High
24	.14	Verbal	Short	Low
36	.13	Verbal	Short	Low
46	.14	Verbal-Visual	Long	Low
50	.11	Verbal	Long	High
56	.14	Verbal	Long	Low
61	-.05	Verbal	Long	Low
64	.01	Verbal	Short	Low

TABLE 11

Type of Discrimination and Correlations

for the Best Ten Items

Item	r_s	Discrimination	Length	Difficulty
18	.43	Verbal-Visual	Long	Low
23	.42	Verbal	Short	Low
41	.40	Verbal-Visual	Short	Low
43	.42	Verbal-Visual	Long	Low
45	.44	Verbal-Visual	Short	Low
58	.47	Verbal	Short	Low
60	.41	Verbal-Visual	Long	Low
80	.42	Verbal-Visual	Short	Low
97	.45	Verbal-Visual	Short	High
100	.40	Verbal-Visual	Short	Low

In order to compare directly the pre- and posttest scores, we selected from the pretest raw data only those items that were administered during the posttest. A new item analysis was performed for both pre- and posttest data with a subsample of 165 subjects who had complete data (both pre- and posttest).

Table 12 presents the Item-Scale correlations and the percentage of subjects answering the item right in each administration as well as the difference (increase or decrease).

The results reported for the pretest data indicated that the type of discriminative clues was an important factor in the stability of the test score. Using the posttest data a similar analysis was carried out. The 55 items were classified according to the characteristics of the stimulus picture. Three groups of items were identified:

a) Items in which the two Gumpgookies look exactly alike, have the same body posture, suggest the same attitude, etc. (Type 1);

b) Items in which the two Gumpgookies are shown in different attitudes but are seen associated (possessing, playing, etc.) with the same objects (Type 2); and

c) Items in which the two Gumpgookies look different (different postures) and are also shown associated with different objects (Type 3). The third column in Table 11 shows the classification of each item.

TABLE 12

Item-Scale Correlations and Percentage Answering Each Item Correctly

Pre- and Posttest 55 Item Form (N = 165)

Item No. Post	Pre	Type of Item	Pretest r.	Posttest r.	Difficulty Pre %	Difficulty Post %	Difference %
1	2	1	.19	.07	68	98	30
2	54	3	.18	.23	70	93	23
3	12	1	.29	.23	63	79	16
4	13	2	.28	.22	84	95	11
5	14	2	.27	.38	88	91	3
6	15	3	.32	.18	85	84	- 1
7	20	1	.33	.16	75	66	- 9
8	21	3	.33	.33	90	93	3
9	22	3	.28	.27	77	78	1
10	25	1	.26	.48	81	81	0
11	28	1	.24	.18	42	87	45
12	30	1	.17	.36	66	83	17
13	31	2	.33	.31	70	88	18
14	32	2	.36	.47	81	89	8
15	33	2	.40	.36	86	90	4
16	37	1	.24	.28	74	88	14
17	38	3	.35	.25	56	62	6
18	40	2	.14	.33	79	81	2
19	41	2	.40	.51	77	81	4
20	44	2	.39	.46	87	89	2

TABLE 12 (Continued)

Item-Scale Correlations and Percentage Answering Each Item Correctly

Pre- and Posttest 55 Item Form (N = 165)

Item No. Post	Pre	Type of Item	Pretest r.	Posttest r.	Difficulty Pre %	Difficulty Post %	Difference %
21	45	2	.41	.44	79	88	9
22	46	2	.19	.31	67	79	12
23	51	1	.22	.40	74	95	21
24	9	2	.01	.50	38	92	54
25	56	1	.19	.26	71	92	21
26	57	1	.33	.24	64	71	7
27	58	1	.47	.46	76	86	10
28	83	2	.33	.42	59	92	33
29	63	1	.42	.18	85	66	- 19
30	64	1	.06	.13	50	57	7
31	65	2	.31	.46	79	78	- 1
32	67	2	.26	.35	66	89	23
33	68	2	.41	.36	83	90	7
34	71	2	.34	.50	80	95	15
35	72	2	.32	.39	64	85	21
36	74	1	.42	.31	61	88	27
37	76	2	.42	.49	81	86	5
38	77	2	.29	.49	80	87	7
39	79	3	.30	.40	45	94	49
40	80	2	.39	.50	84	90	6

TABLE 12 (Continued)

Item-Scale Correlations and Percentage Answering Each Item Correctly

Pre- and Posttest 55 Item Form (N = 165)

Item No. Post	Pre	Type of Item	Pretest r.	Posttest r.	Difficulty Pre %	Difficulty Post %	Difference %
41	81	2	.22	.28	65	70	5
42	82	1	.34	.60	86	93	7
43	59	3	.35	.45	77	79	2
44	87	2	.19	.45	35	83	48
45	88	2	.41	.37	84	94	10
46	90	1	.27	.35	78	75	- 3
47	91	2	.26	.32	85	88	3
48	92	2	.39	.41	79	90	11
49	93	3	.28	.42	83	96	13
50	53	1	.23	.45	67	90	23
51	94	3	.37	.23	85	67	- 18
52	95	1	.32	.54	86	90	4
53	96	2	.37	.34	88	92	4
54	98	1	.35	.46	72	93	21
55	99	3	.29	.28	85	96	11

These three types of items can be thought as constituting a continuum in terms of information provided for the discrimination task and the subsequent choice. At the same time, they can be thought of as ranging from more projective (Type 1) to more objective (Type 3). Thus, when the two Gumpgookies look alike the choice between them has to be made in terms of characteristics ascribed to them on the basis of the verbal description provided by the Examiner. On the other hand, Type 3 items provide more basis for preference and choice in the picture itself, in addition to the verbal instructions provided.

The test was divided into two parts and the mean Item-Scale correlations were compared for the three types of items identified.

Table 13 shows that in general, Type 2 items are better than either Types 1 or 3, and this is more apparent for the first half of the test. Also, in general, the mean Item-Scale correlations are better for the second half than for the first half of the test.

It would seem likely that in general more "projective" items are less effective with young children. On the other hand, these "projective" items are better when they appear later in the testing sequence, suggesting that children "learn" the task better as the testing session progresses. This interpretation is supported by the finding that the Item-Scale correlations for the second half of the test are better than those in the first part.

TABLE 13

Mean Item-Scale Correlations Posttest 55 Item Form

Classified by Discrimination Type

Type	First Half		Second Half		Total
	N	Items 1-27	N	Items 28-55	
1	11	.26	8	.38	.31
2	11	.39	15	.41	.40
3	5	.25	5	.36	.30
Total	27	.32	28	.39	.36

An alternative explanation, suggested by Drs. Adkins and Ballif (personal communication) is that only those highly motivated subjects will finish the test with enough interest to make careful discriminations. In any case, it seems warranted to suggest that Type 1 items should be placed later in the scale in order to maximize this effect.

An examination of those items in which the percentage of Ss answering the item correctly increased from pre- to posttest by 21% or more, revealed the existence of concepts, words and expressions that were probably unfamiliar to the Ss at pretest time and were learned later during the Head Start Program. This would suggest that the test is detecting changes in learning of concepts and expressions. However, this is not inconsistent with the theoretical foundation of the test; Achievement Motivation is learned, and the knowledge of what each alternative implies is no guarantee that the S will choose the motivated Gumpgookie. As a general strategy, though, it seems advisable to eliminate from the text those concepts and expressions which are not a part of most Ss' repertoire.

Summary and Conclusions

The findings reported here are far from complete and conclusive. Analyses carried out in Hawaii by Drs. Adkins and Ballif uncovered major problems in the version of the test administered to our sample. Factor analysis indicated that positional response sets

were contaminating the factors isolated. Thus, a new version of the Gumpgookie was developed in which both the order of the alternatives and the position of the Gumpgookies in the picture are varied. This will hopefully eliminate or decrease significantly the positional response set. It is because of this problem that more detailed comparisons of ethnic groups, sex, ages, etc., were not performed.

However, some of the findings of our analyses are useful in spite of the positional preference problems. Our findings and the solutions suggested have been communicated to Drs. Adkins and Ballif of the Head Start Evaluation and Research Center in Hawaii.